Deployment of Privacy-Preserving Machine Learning for Political Polling in the 2024 Presidential Election

Sam Buxbaum

Lucas M. Tassis, Lucas Boschelli, Giovanni Comarela, Mayank Varia, Mark Crovella, Dino P. Christenson

SysteMPC Workshop

July 10, 2025

Overview

Overview

- We build a system for securely predicting political preferences

Overview

- We build a system for securely predicting political preferences
- We collect and analyze data from almost 8000 unique users

Overview

- We build a system for securely predicting political preferences
- We collect and analyze data from almost 8000 unique users
- All analysis takes place under MPC

Overview

- We build a system for securely predicting political preferences
- We collect and analyze data from almost 8000 unique users
- All analysis takes place under MPC
- Learning algorithm follows a train-update loop until convergence

Overview

- We build a system for securely predicting political preferences
- We collect and analyze data from almost 8000 unique users
- All analysis takes place under MPC
- Learning algorithm follows a train-update loop until convergence
    - Train a logistic regression model on current predictions

Overview

- We build a system for securely predicting political preferences
- We collect and analyze data from almost 8000 unique users
- All analysis takes place under MPC
- Learning algorithm follows a train-update loop until convergence
  - Train a logistic regression model on current predictions
  - Update predictions and repeat

Motivation

Motivation

- Web browsing behavior can predict voting results

Motivation

- Web browsing behavior can predict voting results
- Quantifying the 'Comey letter' (Comarela et al.)

Motivation

- Web browsing behavior can predict voting results
- Quantifying the 'Comey letter' (Comarela et al.)

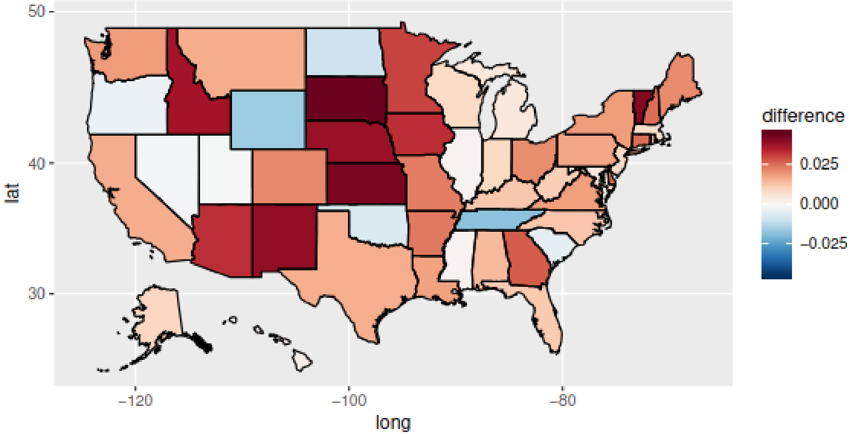


Figure 8: Impact of the 'Comey letter' at the state level.

Motivation

- Web browsing behavior can predict voting results
- Quantifying the 'Comey letter' (Comarela et al.)
  - The event was too close to the election for other polling methods to detect the effect

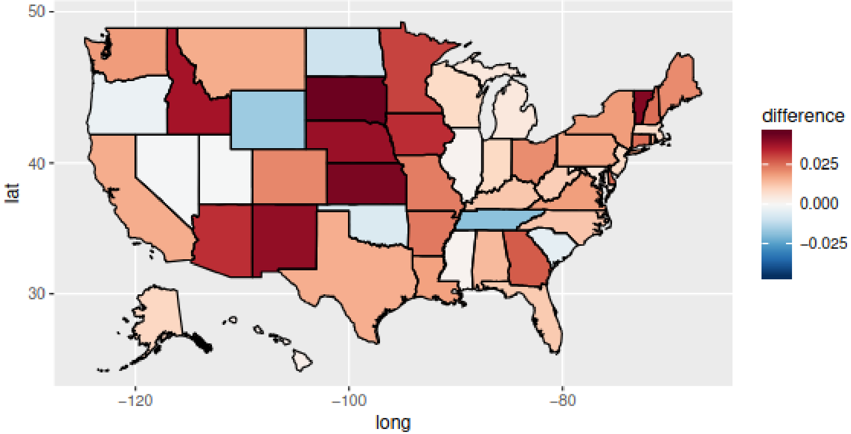


Figure 8: Impact of the 'Comey letter' at the state level.

Two Approaches to Political Polling

Two Approaches to Political Polling

Traditional Polling

- Slow
- Expensive
- Coarse-grained insights

Two Approaches to Political Polling

Traditional Polling	VS	Web Behavior Analysis
Slow		Immediate
Expensive		Cheap
Coarse-grained insights		Fine-grained insights

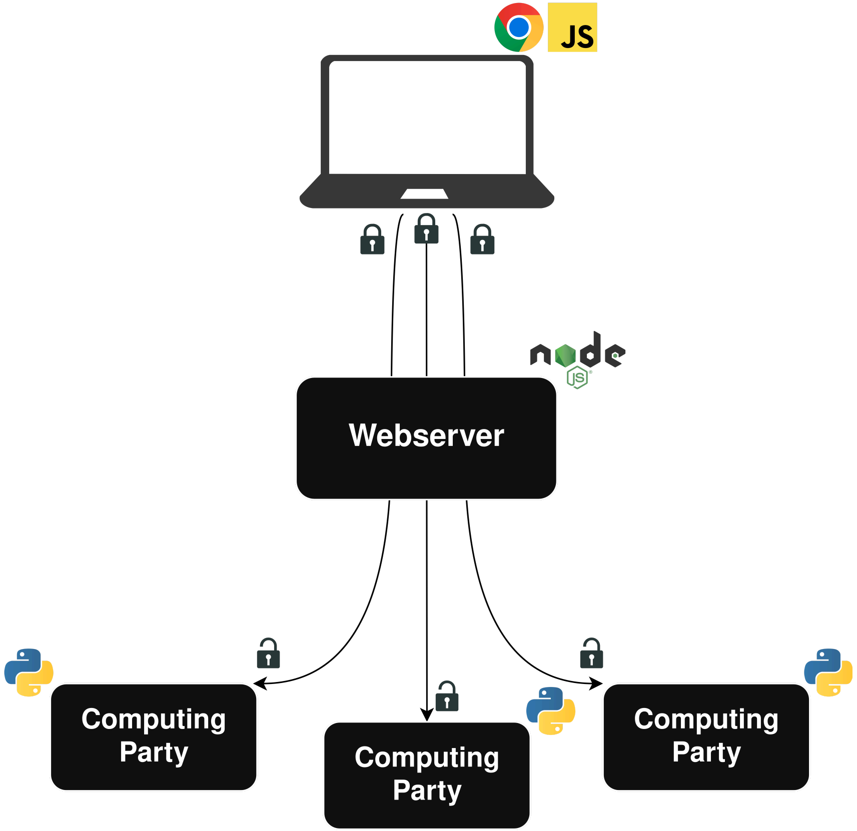
Two Approaches to Political Polling

Traditional Polling	VS	Web Behavior Analysis
Slow		Immediate
Expensive		Cheap
Coarse-grained insights		Fine-grained insights

What about privacy?

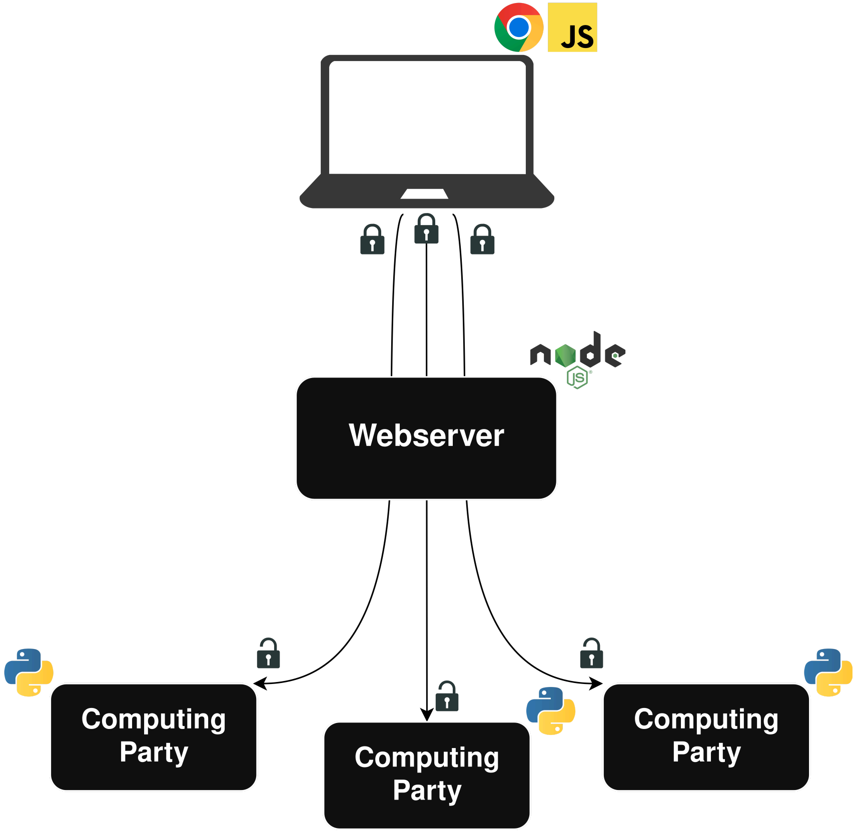
System Design

System Design



System Design

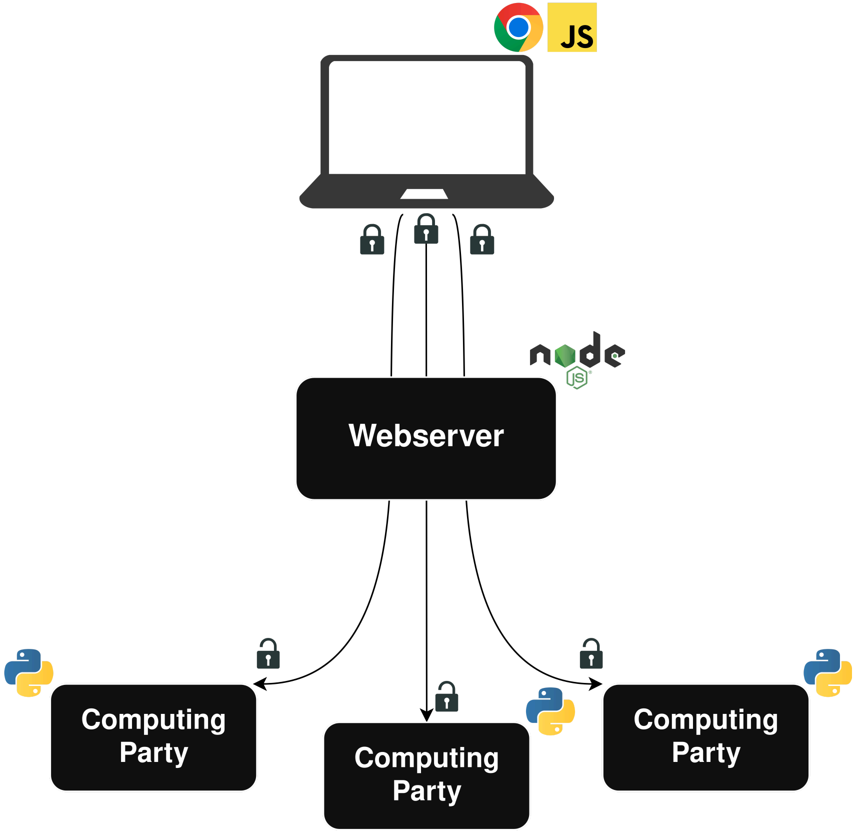
Users



System Design

Users

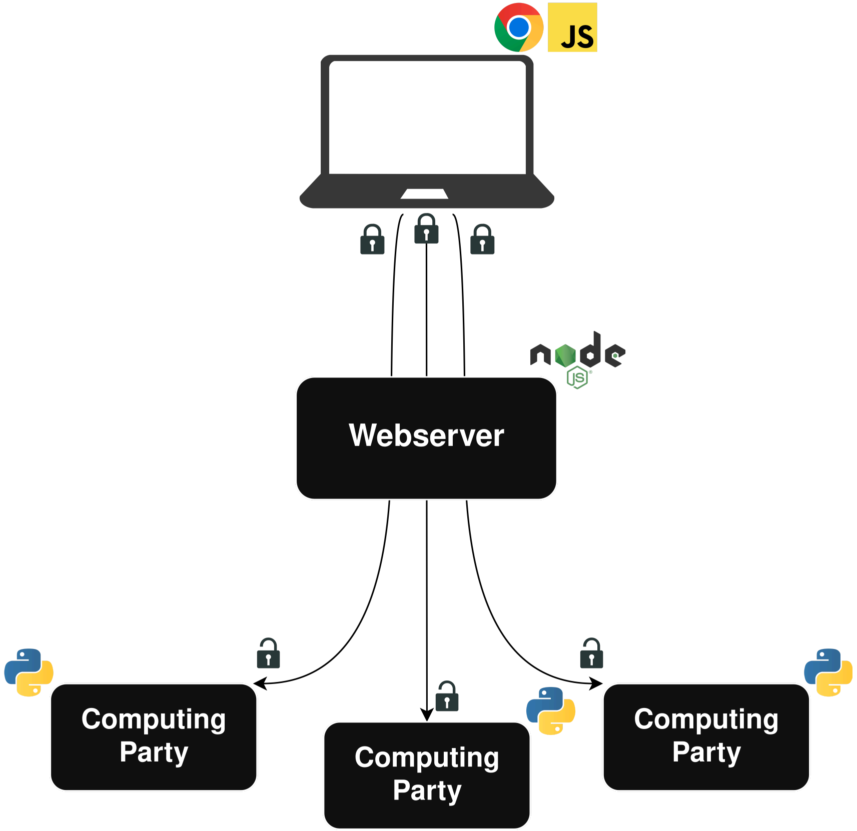
- Built a Chrome plugin to monitor web behavior



System Design

Users

- Built a Chrome plugin to monitor web behavior
- Client-side secret sharing and encryption

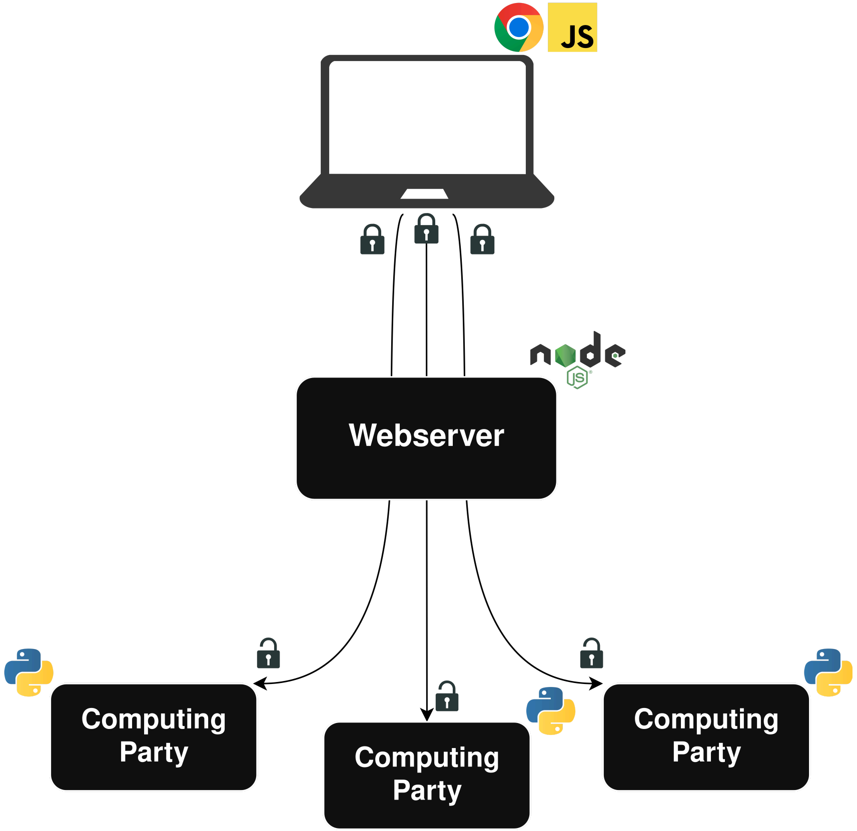


System Design

Users

- Built a Chrome plugin to monitor web behavior
- Client-side secret sharing and encryption

Intermediate webserver



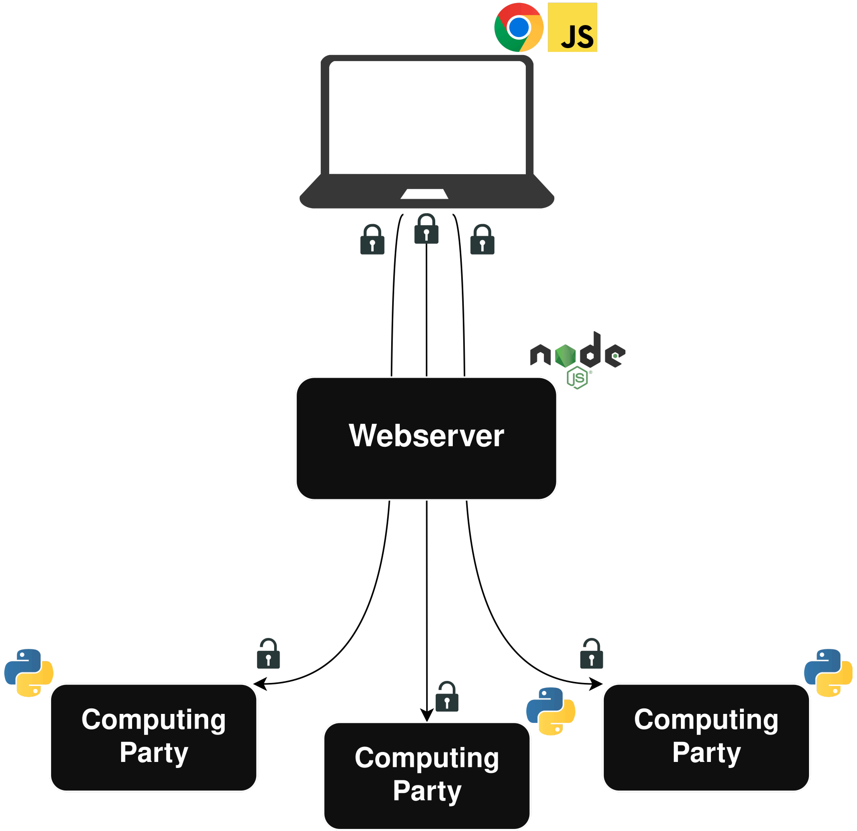
System Design

Users

- Built a Chrome plugin to monitor web behavior
- Client-side secret sharing and encryption

Intermediate webserver

- Simplifies interaction with users



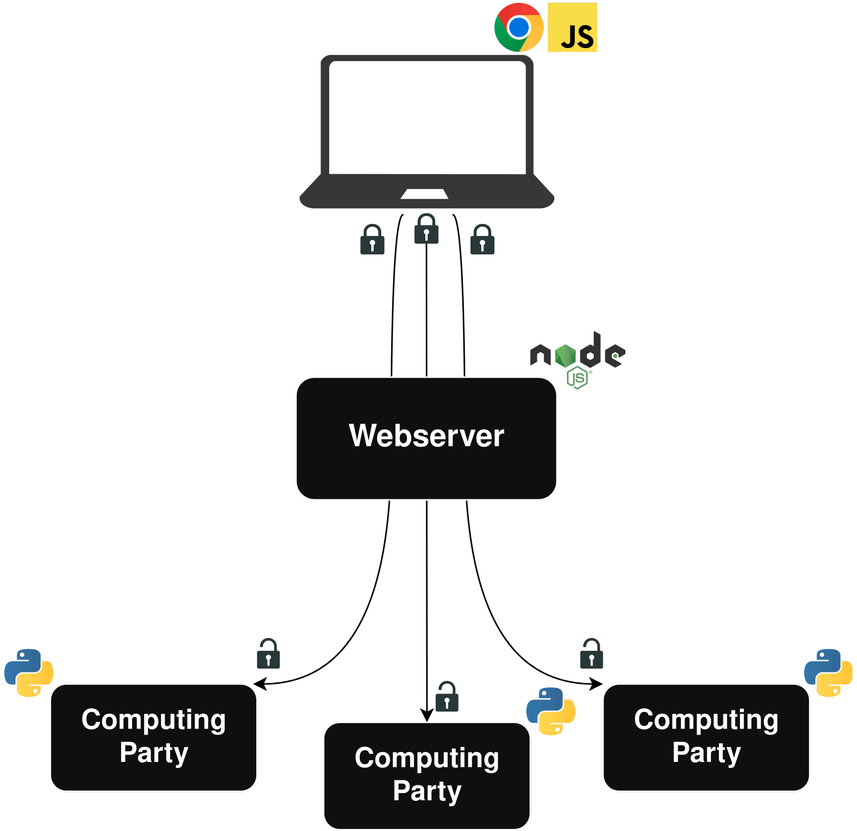
System Design

Users

- Built a Chrome plugin to monitor web behavior
- Client-side secret sharing and encryption

Intermediate webserver

- Simplifies interaction with users
- Collects basic metadata (e.g., for payment)



System Design

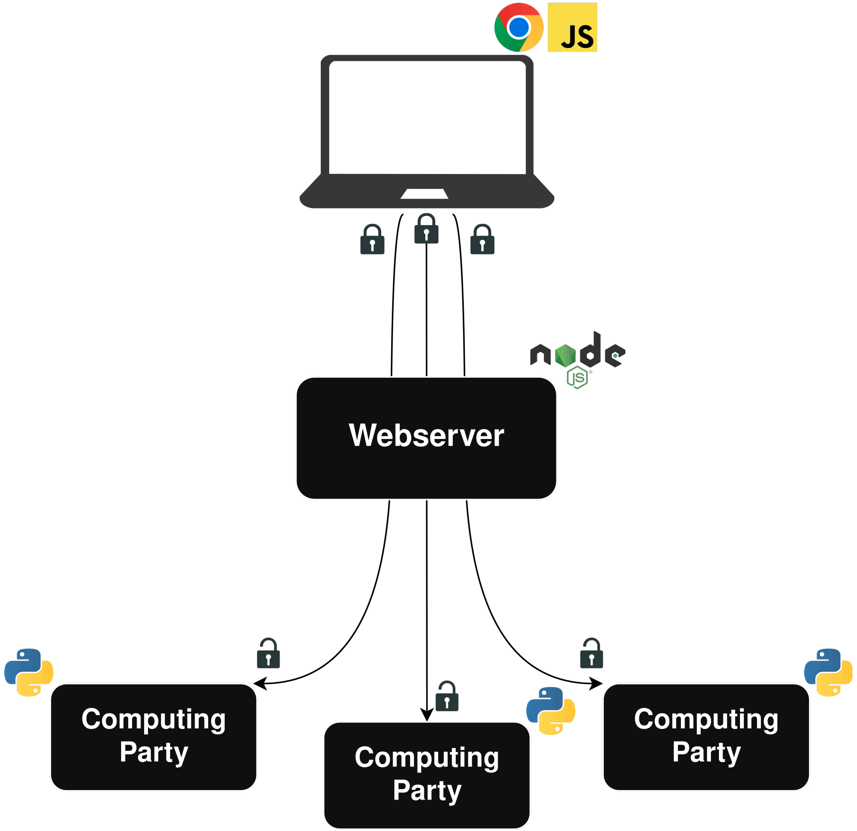
Users

- Built a Chrome plugin to monitor web behavior
- Client-side secret sharing and encryption

Intermediate webserver

- Simplifies interaction with users
- Collects basic metadata (e.g., for payment)

MPC backend



System Design

Users

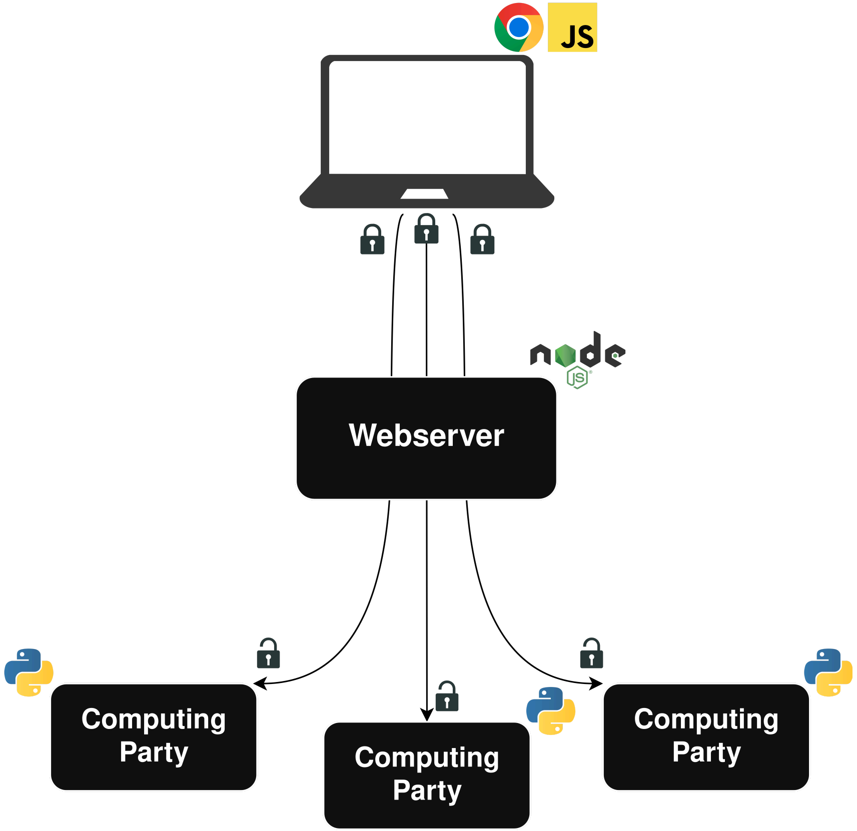
- Built a Chrome plugin to monitor web behavior
- Client-side secret sharing and encryption

Intermediate webserver

- Simplifies interaction with users
- Collects basic metadata (e.g., for payment)

MPC backend

- Trains a model on the data



System Design

Users

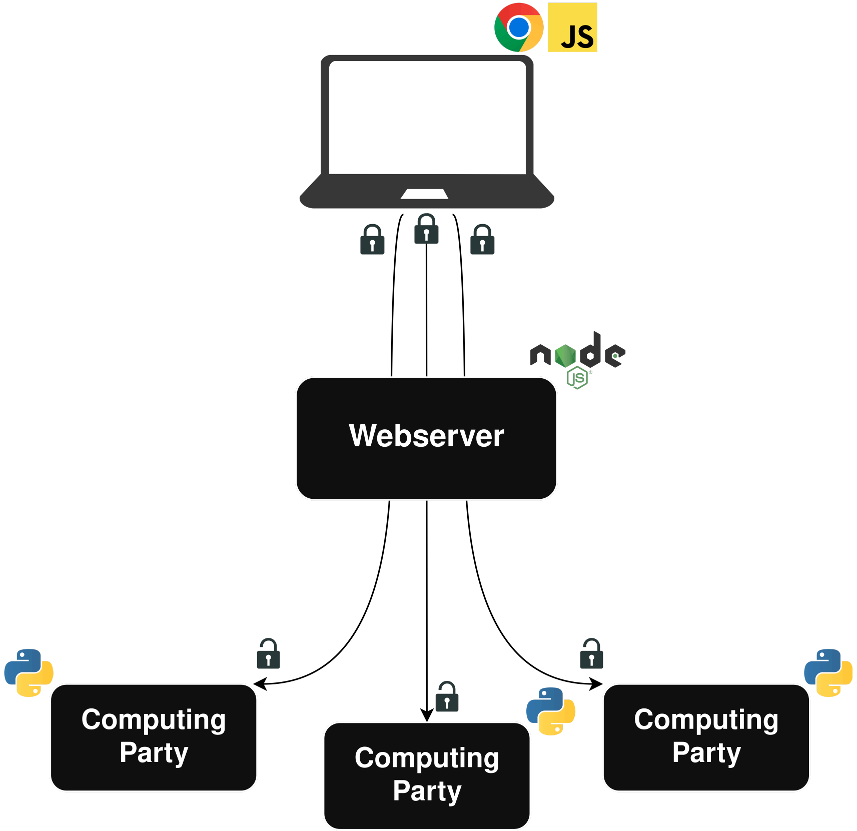
- Built a Chrome plugin to monitor web behavior
- Client-side secret sharing and encryption

Intermediate webserver

- Simplifies interaction with users
- Collects basic metadata (e.g., for payment)

MPC backend

- Trains a model on the data
- We use and augment the CrypTen MPC library



Lessons Learned and Future Directions

Lessons Learned and Future Directions

1. Data integrity matters

Lessons Learned and Future Directions

1. Data integrity matters
   - Verifying user honesty in reporting their state of residence

Lessons Learned and Future Directions

1. Data integrity matters
   - Verifying user honesty in reporting their state of residence
   - How can we balance more extensive tracking with privacy?

Lessons Learned and Future Directions

1. Data integrity matters
    - Verifying user honesty in reporting their state of residence
    - How can we balance more extensive tracking with privacy?

2. Strengthen the threat model

Lessons Learned and Future Directions

1. Data integrity matters
   - Verifying user honesty in reporting their state of residence
   - How can we balance more extensive tracking with privacy?

2. Strengthen the threat model
   - AWS as a single point of trust

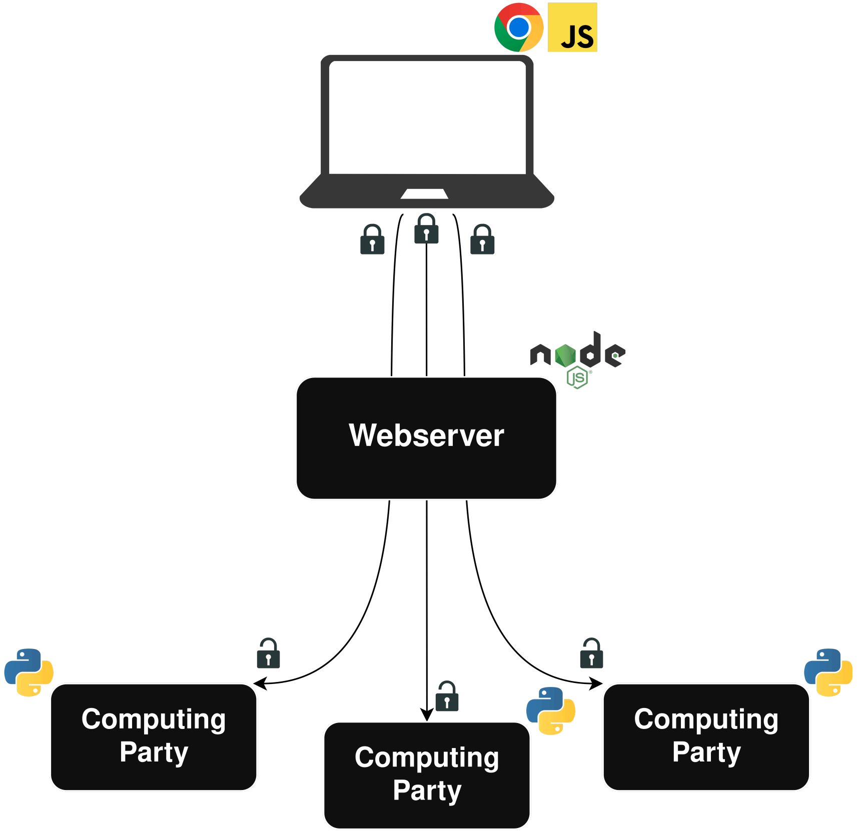
Lessons Learned and Future Directions

1. Data integrity matters
   - Verifying user honesty in reporting their state of residence
   - How can we balance more extensive tracking with privacy?

2. Strengthen the threat model
   - AWS as a single point of trust
   - Anonymous payments

Thank You!

sambux@bu.edu


JS
node JS
Webserver
Computing Party
Computing Party
Computing Party